

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-347917

(43)Date of publication of application : 15.12.2000

(51)Int.Cl.

G06F 12/00

G06F 15/16

G06F 17/30

(21)Application number : 11-156561

(71)Applicant : HITACHI LTD

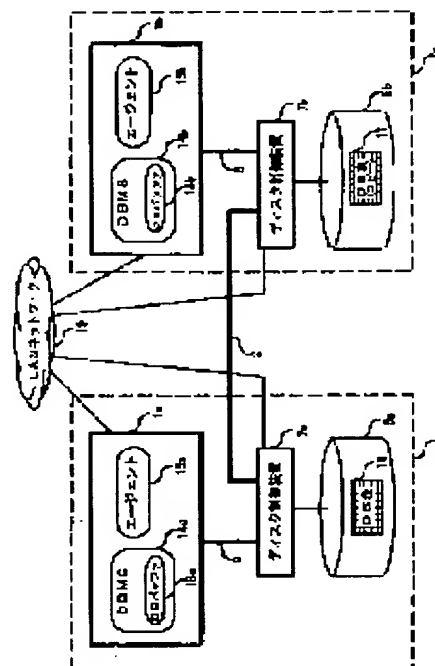
(22)Date of filing : 03.06.1999

(72)Inventor : OBARA KIYOHIRO
TAKAMOTO YOSHIFUMI**(54) METHOD FOR CONTROLLING DATA DUPLEX OPERATION AND DUPLICATED STORAGE SUBSYSTEM**

(57)Abstract:

PROBLEM TO BE SOLVED: To reduce a data transfer time needed to copy DB(database) table data and to guarantee that the data of a copy destination is data having consistency as a DB table.

SOLUTION: A database management system(DBMS) 14a on a host 1a forcedly writes DB table data which is stored in a DB buffer 18a and inputted and outputted by the DBMS 14a to a DB table 16 on a disk drive 5a in response to an instruction of an agent 15a and holds the update of the DB table data even though a subsequent updated request for the DB table data of the table 16 takes place. A disk controller 7a transfers the DB table data of the self-DB table 16 to the DB table on a disk drive 5b through a disk controller 7b, copies the DB table data to duplicate it, interrupts the transfer of subsequent DB table data, makes the DB tables 16 and 17 of the drives 5a and 5b have consistency with each other and enables the DBMS 14b on a host 1b to refer to the copied DB table 17 having consistency.

**LEGAL STATUS**

[Date of request for examination]

27.01.2004

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-347917

(P2000-347917A)

(43) 公開日 平成12年12月15日 (2000. 12. 15)

(51) Int.Cl. ⁷	識別記号	F I	ターミナル (参考)
G 0 6 F 12/00	5 3 1	G 0 6 F 12/00	5 3 1 D 5 B 0 4 5
15/16	6 4 0	15/16	6 4 0 L 5 B 0 7 5
17/30		15/40	3 1 0 C 5 B 0 8 2
		15/401	3 4 0 A

審査請求 未請求 請求項の数 9 O L (全 11 頁)

(21) 出願番号 特願平11-156561

(22) 出願日 平成11年6月3日 (1999. 6. 3)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 小原 清弘

東京都国分寺市東窓ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(72) 発明者 高本 良史

東京都国分寺市東窓ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(74) 代理人 100099298

弁理士 伊藤 修 (外1名)

Fターム (参考) 5B045 DD15 DD18 JJ22 JJ25

5B075 KK03 MM11 ND23 ND30

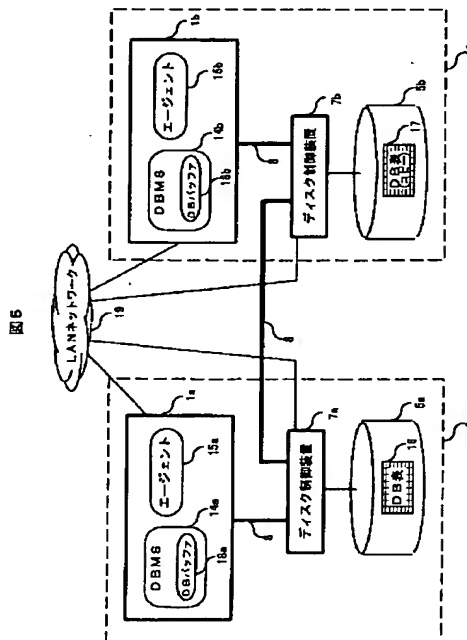
5B082 DE03 DE04 GB07 HA03

(54) 【発明の名称】 データ二重化制御方法および二重化した記憶サブシステム

(57) 【要約】

【課題】 DB表データのコピーに要するデータ転送時間を削減し、コピー先のデータがDB表として整合性の取れたデータであることを保証する。

【解決手段】 ホスト1a上のDBMS 14aは、エージェント15aの指示に応じて、DBバッファ18aに格納されたDBMS 14aで入出力されるデータベース表データ(DB表データ)をディスク装置5a上のデータベース表(DB表)16に強制的に書き出し、該DB表16のDB表データに対する以後の更新要求が起こっても該DB表データの更新を保留し、ディスク制御装置7aは、エージェント15aの指示に応じて、自DB表16のDB表データをディスク制御装置7bを介してディスク装置5b上のDB表に転送してDB表データを複写して二重化し、以後のDB表データの転送を中断し、ディスク装置5a、5bのDB表16、17を整合のとれた状態にし、ホスト1b上のDBMS 14bが整合のとれた複写後のDB表17を参照可能にする。



【特許請求の範囲】

【請求項 1】 第 1 の DBMS が動作する第 1 のプロセッサシステムと該第 1 のプロセッサシステムに接続された第 1 の記憶サブシステムを有する主システムと、第 2 の DBMS が動作する第 2 のプロセッサシステムと該第 2 のプロセッサシステムに接続された第 2 の記憶サブシステムを有する副システムを備え、該第 1 と第 2 の記憶サブシステムが接続されているシステムにおけるデータ二重化制御方法であって、

前記第 1 の DBMS は、前記第 1 のプロセッサシステムの指示に応じて、キャッシュに格納された該第 1 の DBMS で入出力されるデータベースデータ（以下、DB 表データ）を前記第 1 の記憶サブシステムのデータベース表（以下、DB 表）に強制的に書き出し、該 DB 表の DB 表データに対する以後の更新要求が起こっても該 DB 表データの更新を保留し、

前記第 1 の記憶サブシステムは、前記第 1 のプロセッサシステムの指示に応じて、自記憶サブシステムの DB 表の DB 表データを前記第 2 の記憶サブシステムの DB 表に転送して DB 表データを複写して二重化し、以後の DB 表データの転送を中断し、該第 1 と第 2 の記憶サブシステムの DB 表を整合のとれた状態にし、

前記第 2 の DBMS が整合のとれた複写後の DB 表を参照可能にすることを特徴とするデータ二重化制御方法。

【請求項 2】 第 1 の DBMS が動作する第 1 のプロセッサシステムと該第 1 のプロセッサシステムに接続された第 1 の記憶サブシステムを有する主システムと、第 2 の DBMS が動作する第 2 のプロセッサシステムと該第 2 のプロセッサシステムに接続された第 2 の記憶サブシステムを有する副システムを備え、該第 1 と第 2 の記憶サブシステムが接続されているシステムにおけるデータ二重化制御方法であって、

前記第 1 の DBMS は、前記第 1 の記憶サブシステムの指示に応じて、キャッシュに格納された該第 1 の DBMS で入出力される DB 表データを該第 1 の記憶サブシステムの DB 表に強制的に書き出し、該 DB 表の DB 表データに対する以後の更新要求が起こっても該 DB 表データの更新を保留し、

前記第 1 の記憶サブシステムは、自記憶サブシステムの DB 表の DB 表データを前記第 2 の記憶サブシステムの DB 表に転送して DB 表データを複写して二重化し、以後の DB 表データの転送を中断し、該第 1 と第 2 の記憶サブシステムの DB 表を整合のとれた状態にし、

前記第 2 の DBMS が整合のとれた複写後の DB 表を参照可能にすることを特徴とするデータ二重化制御方法。

【請求項 3】 請求項 1 または請求項 2 記載のデータ二重化制御方法において、

前記第 1 の記憶サブシステムの DB 表に強制的に書き出した DB 表データを該第 1 の記憶サブシステムのリモートコピー用バッファに格納し、

該第 1 の記憶サブシステムは、二重化の指示に応じて該バッファ内の DB 表データを前記第 2 の記憶サブシステムの DB 表に転送して DB 表データを複写して二重化することを特徴とするデータ二重化制御方法。

【請求項 4】 請求項 1 または請求項 2 記載のデータ二重化制御方法において、

前記第 1 の記憶サブシステムの DB 表に強制的に書き出した DB 表データの該 DB 表における位置情報を該第 1 の記憶サブシステム内のメモリに記憶しておき、

該第 1 の記憶サブシステムは、二重化の指示に応じて該メモリ内の位置情報を読み出し、該位置情報の示す DB 表の位置に記憶されている DB 表データを前記第 2 の記憶サブシステムの DB 表に転送して DB 表データを複写して二重化することを特徴とするデータ二重化制御方法。

【請求項 5】 第 1 のプロセッサシステムに接続され、第 1 の外部記憶制御装置と第 1 の外部記憶装置を有する第 1 の記憶サブシステムと、第 2 のプロセッサシステムに接続され、第 2 の外部記憶制御装置と第 2 の外部記憶装置を有する第 2 の記憶サブシステムを備え、該第 1 と第 2 の外部記憶制御装置が接続された二重化した記憶サブシステムであって、

前記第 1 の外部記憶制御装置は、前記第 1 のプロセッサシステム内のキャッシュに記憶されたデータを前記第 1 の外部記憶装置に強制的に書き出す手段と、該第 1 の外部記憶装置に書き込まれたデータに対する以後の更新要求が起こっても該データの更新を保留する手段と、該データを前記第 2 の記憶サブシステムに転送して該データを二重化する手段と、該データの該第 2 の記憶サブシステムへの転送後、以後の該第 1 の外部記憶装置から該第 2 の記憶サブシステムへのデータの転送を中断する手段を有し、

該第 1 と第 2 の記憶サブシステムに記憶されたデータを整合のとれた状態にすることを特徴とする二重化した記憶サブシステム。

【請求項 6】 請求項 5 記載の二重化した記憶サブシステムにおいて、

前記第 1 の外部記憶制御装置は、前記第 1 の外部記憶装置に強制的に書き出したデータを第 1 の外部記憶制御装置内のバッファに格納する手段と、二重化の指示に応じて該バッファ内のデータを前記第 2 の記憶サブシステムに転送して該データを二重化する手段を有することを特徴とする二重化した記憶サブシステム。

【請求項 7】 請求項 5 記載の二重化した記憶サブシステムにおいて、

前記第 1 の外部記憶制御装置は、前記第 1 の外部記憶装置に強制的に書き出したデータの該第 1 の外部記憶装置における位置情報を第 1 の外部記憶制御装置のメモリに記憶する手段と、二重化の指示に応じて該メモリ内の位置情報を読み出し、該位置情報の示す該第 1 の外部記憶

装置の位置に記憶されているデータを前記第2の記憶サブシステムに転送して該データを二重化する手段を有することを特徴とする二重化した記憶サブシステム。

【請求項8】 請求項5乃至請求項7のいずれかの請求項記載の二重化した記憶サブシステムにおいて、前記第1の外部記憶制御装置と第2の外部記憶制御装置の間を専用の通信線で結合し、データ通信を行うようにすることを特徴とする二重化した記憶サブシステム。

【請求項9】 請求項5乃至請求項7のいずれかの請求項記載の二重化した記憶サブシステムにおいて、前記第1の外部記憶制御装置と第2の外部記憶制御装置の間をスイッチまたはネットワークで結合し、データ通信を行うようにすることを特徴とする二重化した記憶サブシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はデータの二重化に係り、特に、DBMSのデータに対するデータ二重化制御方法およびデータ二重化のための二重化した記憶サブシステムに関する。

【0002】

【従来の技術】最近、データウェアハウスシステムと呼ばれる、業務上データ解析システムが普及してきている。データウェアハウスはデータベース管理システム(DBMS)であり、そのデータベースの元データは、一般に基幹業務用データベースからデータを抽出して作成している。またデータウェアハウスシステムは、大量のデータに対し多角的な解析を行うため、多量のCPUパワーを消費する。このため、基幹業務システムとは別のシステムで構築され、LAN(ローカルエリアネットワーク)やWAN(ワイドエリアネットワーク)を用いてデータ通信可能な構成になっている。一般に基幹業務システムからデータウェアハウスシステムへのデータのロードは次のような手順で行われている。

(1) 基幹業務システム上で、基幹業務用DBMSが、必要なデータの抽出。

(2) 上記抽出データを、FTP(ファイル転送プログラム)等により、データウェアハウスシステムへ転送。

(3) データウェアハウスシステム上で、上記転送されたデータを、データウェアハウス用データベースへロード。

上記手順において、(2)で転送される抽出データの大きさは、大規模な企業では数十ギガバイト以上にもなる。そしてこのデータ転送は、一般にLAN/WANを用いて行っていたため、データ転送に非常に時間がかかる。さらに転送中はLAN/WANのレスポンスが下がり、他の目的でLAN/WANを利用しているユーザにまで影響が及ぶという問題がある。また(1)のデータ抽出においても、データ抽出のために基幹業務システムのCPUパワーが消費され、本来の基幹業務のレスポ

スが悪化するという問題点もある。

【0003】このデータ転送の問題点は、リモートコピー等、ディスクシステムが自動的にデータの二重化を行う方式を用いて、予め事前に基幹業務のデータをデータウェアハウスに送っておくことにより解決されるように見える。すなわち基幹業務データベース表(DB表)のコピーを、データウェアハウスシステム側に事前に生成しておくのである。これにより、上記手順の(2)が省略でき、下記に示すような手順でデータウェアハウスへのデータロードが実行できる。

(1) データウェアハウスシステム上で、コピーされた基幹業務データベース表データから、必要なデータを抽出。

(2) データウェアハウスシステム上で、上記抽出されたデータを、データウェアハウス用データベースへロード。

【0004】

【発明が解決しようとする課題】前述した方式は、データ抽出処理もデータウェアハウス側に移管され、基幹業務システムの負荷を減らすという点でも良い選択肢のように見える。しかしデータベース表データの整合性に関し、新たな問題が発生する。DBMSは、自身内にデータベース表データの一部を保持するキャッシュ(DBバッファ)を持っている。このため、DB表データの更新が起こっても、直ぐにディスク上のDB表データへ反映されない。また更新されるタイミングも任意である。すなわち、任意の時点では基幹業務システムのディスクシステム上のDB表データは整合性が取れた状態とは限らないため、それをコピーしたデータも整合性が取れた状態とは限らない。本発明の目的は、データをFTP等のファイル転送で送る代わりに、リモートコピー等ディスクシステムが自動的に事前にデータの二重化を行う方式を用いて、DB表データをコピーしておき、データ転送時間を削減する方式において、コピー先のデータがデータベース表として整合性の取れたデータであることを保証する手段を提供することにある。これにより、データの転送時間を削減すると共に、基幹業務システム上でデータ抽出処理を、データウェアハウスシステム側にオフロードでき、データ抽出処理時における基幹業務システムのレスポンス低下も回避できる。

【0005】

【課題を解決するための手段】上記目的を達成するため、本発明は、第1のDBMSが動作する第1のプロセッサシステムと該第1のプロセッサシステムに接続された第1の記憶サブシステムを有する主システムと、第2のDBMSが動作する第2のプロセッサシステムと該第2のプロセッサシステムに接続された第2の記憶サブシステムを有する副システムを備え、該第1と第2の記憶サブシステムが接続されているシステムにおけるデータ二重化制御方法であり、前記第1のDBMSは、前記第

10

20

30

40

50

1のプロセッサシステムの指示に応じて、キャッシュに格納された該第1のDBMSで入出力されるデータベースデータ（以下、DB表データ）を前記第1の記憶サブシステムのデータベース表（以下、DB表）に強制的に書き出し、該DB表のDB表データに対する以後の更新要求が起こっても該DB表データの更新を保留し、前記第1の記憶サブシステムは、前記第1のプロセッサシステムの指示に応じて、自記憶サブシステムのDB表のDB表データを前記第2の記憶サブシステムのDB表に転送してDB表データを複写して二重化し、以後のDB表データの転送を中断し、該第1と第2の記憶サブシステムのDB表を整合のとれた状態にし、前記第2のDBMSが整合のとれた複写後のDB表を参照可能にするようにしている。

【0006】また、第1のDBMSが動作する第1のプロセッサシステムと該第1のプロセッサシステムに接続された第1の記憶サブシステムを有する主システムと、第2のDBMSが動作する第2のプロセッサシステムと該第2のプロセッサシステムに接続された第2の記憶サブシステムを有する副システムを備え、該第1と第2の記憶サブシステムが接続されているシステムにおけるデータ二重化制御方法であり、前記第1のDBMSは、前記第1の記憶サブシステムの指示に応じて、キャッシュに格納された該第1のDBMSで入出力されるDB表データを該第1の記憶サブシステムのDB表に強制的に書き出し、該DB表のDB表データに対する以後の更新要求が起こっても該DB表データの更新を保留し、前記第1の記憶サブシステムは、自記憶サブシステムのDB表のDB表データを前記第2の記憶サブシステムのDB表に転送してDB表データを複写して二重化し、以後のDB表データの転送を中断し、該第1と第2の記憶サブシステムのDB表を整合のとれた状態にし、前記第2のDBMSが整合のとれた複写後のDB表を参照可能にするようにしている。

【0007】また、前記第1の記憶サブシステムのDB表に強制的に書き出したDB表データを該第1の記憶サブシステムのリモートコピー用バッファに格納し、該第1の記憶サブシステムは、二重化の指示に応じて該バッファ内のDB表データを前記第2の記憶サブシステムのDB表に転送してDB表データを複写して二重化するようにしている。

【0008】また、前記第1の記憶サブシステムのDB表に強制的に書き出したDB表データの該DB表における位置情報を該第1の記憶サブシステム内のメモリに記憶しておき、該第1の記憶サブシステムは、二重化の指示に応じて該メモリ内の位置情報を読み出し、該位置情報の示すDB表の位置に記憶されているDB表データを前記第2の記憶サブシステムのDB表に転送してDB表データを複写して二重化するようにしている。

【0009】また、第1のプロセッサシステムに接続さ

れ、第1の外部記憶制御装置と第1の外部記憶装置を有する第1の記憶サブシステムと、第2のプロセッサシステムに接続され、第2の外部記憶制御装置と第2の外部記憶装置を有する第2の記憶サブシステムを備え、該第1と第2の外部記憶制御装置が接続された二重化した記憶サブシステムであり、前記第1の外部記憶制御装置は、前記第1のプロセッサシステム内のキャッシュに記憶されたデータを前記第1の外部記憶装置に強制的に書き出す手段と、該第1の外部記憶装置に書き込まれたデータに対する以後の更新要求が起こっても該データの更新を保留する手段と、該データを前記第2の記憶サブシステムに転送して該データを二重化する手段と、該データの該第2の記憶サブシステムへの転送後、以後の該第1の外部記憶装置から該第2の記憶サブシステムへのデータの転送を中断する手段を有し、該第1と第2の記憶サブシステムに記憶されたデータを整合のとれた状態にするようにしている。

【0010】また、前記第1の外部記憶制御装置は、前記第1の外部記憶装置に強制的に書き出したデータを第1の外部記憶制御装置内のバッファに格納する手段と、二重化の指示に応じて該バッファ内のデータを前記第2の記憶サブシステムに転送して該データを二重化する手段を有するようにしている。

【0011】また、前記第1の外部記憶制御装置は、前記第1の外部記憶装置に強制的に書き出したデータの該第1の外部記憶装置における位置情報を第1の外部記憶制御装置のメモリに記憶する手段と、二重化の指示に応じて該メモリ内の位置情報を読み出し、該位置情報の示す該第1の外部記憶装置の位置に記憶されているデータを前記第2の記憶サブシステムに転送して該データを二重化する手段を有するようにしている。

【0012】また、前記第1の外部記憶制御装置と第2の外部記憶制御装置の間を専用の通信線で結合し、データ通信を行うようにしている。

【0013】また、前記第1の外部記憶制御装置と第2の外部記憶制御装置の間をスイッチまたはネットワークで結合し、データ通信を行うようにしている。

【0014】

【発明の実施の形態】次に図1～図6を用いて本発明の一実施例を説明する。図2は、ホストコンピュータに接続された単一のディスク制御装置の一構成例を示す。ディスク制御装置7は、ホストインタフェース2とチャネルバス8によりホストコンピュータ1と接続し、キャッシュメモリ3、共有メモリ20、ディスク駆動装置5と接続するディスクインタフェース4およびこれらを接続する共通バス6により構成されている。複数のディスクインタフェースが実装可能である。複数のホストインタフェース2も実装可能であり、この場合接続先のホスト1が同一であるか否かは問わない。本実施例の図2では、各ホストインタフェースが同じホストに接続されて

いる例を示している。ホストインタフェース 2 とディスクインタフェース 4 には、プロセッサが装備されており、それぞれ自律的に動作する。またキャッシュメモリ 3、LAN インタフェース 11 は、複数のホストインタフェース 2 や複数のディスクインタフェース 4 から参照可能な共有資源である。キャッシュメモリ 3 には、本ディスク制御装置へ書き込まれたデータやディスク駆動装置 5 から読み出されホストに出力したデータが一時的に保存される。ディスク制御装置がディスクアレイ機能を持っている場合には、ホストから送られたデータは分割され、複数のディスク駆動装置 5 に分散して格納される。本実施例はディスクアレイに対しても適用可能であるが、説明の簡略化のため、以後の説明は通常のディスクに対する動作を例に説明する。

【0015】図 3 を用いてリモートコピーシステムの構成例を示す。リモートコピーは、ディスク制御装置が自律的に、指定されたボリュームを他のディスク制御装置にコピーする機能である。具体的は、ホストインタフェース上のプログラムにより実現される。図 3 において、9 は主システムであり、ホストコンピュータ 1 a と、ディスク制御装置 7 a と、ディスク駆動装置 5 a を有し、2 はホストインタフェース 2 であり、8 はチャネルバスである。10 は副システムであり、主システムと同様に、ホストコンピュータ 1 b と、ディスク制御装置 7 b と、ディスク駆動装置 5 b を有し、2 はホストインタフェース 2 であり、8 はチャネルバスである。ここでは、主システム 9 のディスク駆動装置 5 a 中のボリューム A を、副システム 10 のディスク駆動装置 5 b 中へコピーしている。図 3 では、主システムと副システムが同一の構成のように見えるが、リモートコピーを行うに当たり、一般には主システムと副システムが、稼働しているソフトウェアも含め同一構成システムである必要はない。さらに言えば、ここでは便宜的に主システム／副システムと呼んでいるが、一方が他方の待機系である必要もない。例えば主システムが基幹業務システムであり、副システムがデータウェアハウスシステムでも良い。図 3 においても、ボリューム A 以外のボリュームは、異なるアプリケーションが利用している異なる内容のボリュームを仮定している。

【0016】リモートコピーの動作は、次のようになる。まず、ホスト 1 a からディスクへの書込要求コマンドに対し、ホストインタフェース #0 は、その書込先のボリュームがリモートコピーの対象になっているか否かを判断する。リモートコピーの対象になっていない場合は、そのまま書込要求コマンドを処理する。書込先がリモートコピーの対象になっている場合は、通常通りに書込要求コマンドを処理すると共に、副システム 10 のディスク制御装置 7 b と接続されているホストインタフェース #1 を使用して、ホストから受け取ったコマンドと同一の書込要求コマンドをディスク制御装置 7 b に発行

する。これにより、副システム 10 のディスク駆動装置 5 b 上に、ボリューム A の複製が生成される。ここでホストインタフェースは、入出力コマンドの発行処理と受信処理の両方の機能を備えている。これらの機能はホストインタフェース中のプロセッサの処理により実現されている。リモートコピーは、コピー先への書込コマンドの発行方法の違いにより、同期リモートコピーと非同期リモートコピーに分類できる。同期リモートコピーは、主システム 9 のディスク制御装置 7 a がホスト 1 a から書込要求コマンドを受け取った後、副システム 10 のディスク制御装置 7 b に同一の書込要求コマンドを発行し、その確認応答が返された後に、ホスト 1 a に確認応答を返す。一方非同期リモートコピーは、その名が示すとおり、ホスト 1 a への確認応答を返した後の適当なタイミングでディスク制御装置 7 b へ書込要求コマンドを発行する。

【0017】このような処理方法の違いにより、同期リモートコピーと非同期リモートコピーは次のような特徴の差異が生じる。まず、ホスト 1 a から見た書込コマンドの応答時間については、同期リモートコピーは二回の書込コマンドを処理するため、応答時間が従来比約 2 倍伸びるのに対し、非同期リモートコピーは従来とほぼ同じである。また非同期リモートコピーは、ディスク制御装置 7 b への書込要求コマンドを非同期で発行するため、リモートコピーするデータを一時保存しておくバッファが必要となる。このバッファにはキャッシュメモリが利用される。

【0018】リモートコピーの開始／終了等は、通常の入出力命令と同様なコマンドを用いホスト上のプログラムから制御される。主なコマンドを次に述べる。

(1) 初期化&コピー開始コマンド（コピー先のボリュームの内容をコピー元と同一にするために指定されたボリュームの全内容をコピー先へコピーする（初期化）と共に、ホストから発行された書込要求コマンドに対し、指定されたコピーモード（同期／非同期）でリモートコピーを開始する）。

(2) 中断コマンド（リモートコピーを一時中断する。この後に受け付けた書込要求コマンドに対するリモートコピーデータはバッファに保存しておき、後の再開コマンドに備える）。

(3) 再開コマンド（中断していたリモートコピーを再開する。バッファに保存されているリモートコピーデータに対してもコピーを行う）。

(4) フラッシュコマンド（バッファに保存されているリモートコピーデータを強制的にコピー先へコピーする）。

【0019】図 4 にデータベース管理システム（DBMS）14 とディスク内の DB 表 16 およびログファイル 21 の関係を示す。DB 表はデータベースのデータである。DBMS は DB 表をデータベースバッファ（DB バ

ッファ) 18を通して読み込む。DBバッファはいわゆるキャッシュであり、最近使用されたデータを格納することにより、ディスクアクセス回数を削減し処理の高速化を図る。ログファイル21は、DBバッファ18やDB表16の変更内容、コミットの情報等、データベースの障害回復に必要な情報が格納される。これらDBMSの機能や処理の詳細に関しては、例えばオーム社発行の「データベースシステムとデータモデル」穂鷹良介著に述べられている。

【0020】ところで、DBバッファ18のキャッシュとしての制御は、いわゆるライトバック方式である。すなわち、更新されたデータはキャッシュが一杯になるまで保存され、直ちにディスク上のDB表16には書き込まれない。したがってディスク上のDB表16を見た場合、そのデータは最新データであるとは限らないし、更新されるタイミングも不定である。これではこのDB表をリモートコピーし、コピー先で参照しても、整合の取れた状態であるとは限らない。整合の取れた状態とは、任意の時点での最新状態である事と、そのDB表を参照中に内容が変更されない事が保証されている状態である。

【0021】ここで、ディスク上のDB表16を整合の取れた状態にし、リモートコピー先で安全に参照できるように、DBMS14における処理として次の二つの機能を加える。一つは、DBバッファ18の内容を強制的にディスク上のDB表16に書き出す機能である。これをDBバッファのフラッシュと呼ぶ。もう一つは、DBバッファが一杯になっても、その内容をディスク上のDB表16に書き込まないようにする機能である。これをDB表への書込抑止と呼ぶ。これらの機能は、すべてDBMS14のソフトウェアによって容易に実現される。例えばDB表16への書込抑止機能は、ログファイル21を利用することにより実現される。ディスク上のDB表16への書込を単純に抑制しても、ログファイル21には変更すべき内容を全て記録するため、DB表16の変更内容が失われることはない。そして一時的にディスク上のDB表16の内容の更新を抑止しても、コピー元のDBMS側では、任意の時点でログファイル21の内容を参照することにより、正常に内容の更新が行われた場合と同じ状態に復元できる。

【0022】これら、DBバッファのフラッシュ、DB表への書込抑止、DB表への書込再開等の指示は、DBMSへの通常の検索問い合わせと同様な方式、いわゆるSQLコマンドによって提供される。したがって、ホスト上やホストとLAN等で接続された任意のシステムから、これらの機能を利用及び制御することが出来る。

【0023】図5はホストコンピュータにDBMSを備えている場合のデータの二重化を行うシステムを示す。図5において、9は主システム、10は副システムであり、1a、2aはそれぞれ主、副のホストコンピュー

タ、7a、7bはそれぞれ主、副のディスク制御装置、5a、5bはそれぞれ主、副のディスク駆動装置、14a、14bはそれぞれ主、副のDBMS、15a、15bはそれぞれ主、副のエージェント、18a、18bはそれぞれ主、副のDBバッファ、16、17はそれぞれDB表、8はチャンネルバスである。

【0024】図5のシステム構成に対し、整合の取れたDB表をリモートコピーし、副システム側で参照する手順を図1に示す。ここでエージェント15a、15bとは、それぞれのホスト上で動作するプログラムであり、DBMS14a、14bやディスク制御装置7a、7bの適切な制御を行う。まず、主ホスト上エージェント15aは、DBMS14aに対し、DBバッファ中データのディスクへのフラッシュのコマンドを発行し、次いで、ディスク上のDB表データ16への書込を抑止するコマンドを発行し、DB表16を整合の取れた状態にする。これらの手続きはDBMS16をいわゆるバックアップモードに移行する手続きと同様である。バックアップはデータをレストアした場合を考慮し、整合の取れた状態で行われるのが一般的だからである。次に、リモートコピーの転送元のDB表16（以下、主DB表16）と転送先のDB表17（以下、副DB表17）の内容を一致させるために、主ディスク制御装置7a内のリモートコピー用のバッファに入っているデータがあれば、強制的にリモート側へ転送させる。これは非同期リモートコピーの場合にだけ必要である。同期リモートコピーの場合は、主システムのディスク制御装置7aで書き込まれた内容は直ちに副システムへ反映されるので必要ない。そして、リモートコピー終了後、以後のリモートコピーを中断させる。これで、主システム中の主DB表16が更新されても、その影響が副システムの副DB表17に及ばないようにする。これにより、副システム中のDBMS14bは任意の時点のDB表の内容を参照可能となる。また引き続き、DBMS14aを通常モードに移行、すなわちDB表データ16への書込を再開するコマンドを発行する。この再開コマンドは必須ではない。しかし、あまり長い間主DB表16の書込を抑止していると、ログファイルを利用して主DB表16の内容を抑止が無かった状態に復元する時に、復元する箇所が多くなり復元に時間がかかる。

【0025】さらに副ホスト上エージェント15bに対し、LANネットワーク19を通じて動作開始を指示する。副ホスト上エージェント15bは、DBMS14bに対し、副DB表17をオンラインにする事を指示する。これによりDBMS14bは副DB表17を利用することが可能になる。この後、副ホスト上エージェント15bは、DBMS14bを通じて副DB表17を参照する任意のプログラムを起動、もしくは準備が整った事を通知することにより、本発明の目的が達せられる。ここで、エージェントは主システム副システムそれぞれに

存在する必要は無い。例えば主システムだけに存在していても本発明の目的は達せられる。この場合副システムのDBMS14bへの指示等は、LANネットワーク19を通じて行う。

【0026】また同様に、エージェントの機能がディスク制御装置に内蔵されている構成でも本発明の目的は達せられる。この場合エージェントプログラムを実行するのは、ディスク制御装置内のホストインタフェースのプロセッサである。ディスク制御装置もLANインタフェース11を持っているため、LANネットワークを通してホストと他のディスク制御装置と通信及び処理の同期ができる。この構成における手順を図6に示す。

【0027】図3および図5において、ディスク制御装置7a、7b間はチャンネルバス8のような専用の通信線で接続されリモートコピーを行っている構成を説明しているが、図7のように、ホストおよびディスク制御装置間をスイッチやネットワークで結合し、リモートコピーやデータの二重化を行う構成に対しても本発明は適用可能である。

【0028】これまでの実施例では、非同期リモートコピーに関して、ディスク制御装置内のキャッシュメモリを、リモート側へデータを書き込むためデータを一時保存しておくバッファとして利用していると述べた。しかし、データのバッファリングに関しては、これ以外にも次に述べるようなやり方がある。それは、書き込まれたデータそのものをキャッシュにバッファリングするのではなく、書き込まれたデータのディスク上の位置（ボリューム番号、シリンダ番号、ヘッド番号、レコード番号等）の情報をログとして記憶するものである。すなわち、リモートコピー機能が有効になっていてもコピー元のディスク制御装置のボリュームにデータが書かれることを利用し、キャッシュ上ではなくディスクドライブ上に保存しておくものである。

【0029】このやり方では、図8に示すような書き込みデータのログ情報を使用する。このログは共有メモリ上に置かれ、ホストからの書き込み要求を受けたホストインタフェースによって各エントリが生成される。図8の最初の4つのエントリは、（ボリューム番号0、シリンダ番号1、ヘッド番号1、レコード番号1）、（ボリューム番号0、シリンダ番号1、ヘッド番号1、レコード番号2）、（ボリューム番号0、シリンダ番号1、ヘッド番号1、レコード番号3）、（ボリューム番号0、シリンダ番号5、ヘッド番号1、レコード番号1）、の4個のデータが更新されたことを示す。リモートと接続されたホストインタフェースは、このログを参照して書き込まれたデータの位置の情報を得ると共に、そのデータをリモート側へコピーする。コピーが完了したログのエントリは、リモートと接続されたホストインタフェースにより削除される。このようなログを使用することにより、キャッシュ上でのバッファリングより少ない情報量で、リ

モートへコピーすべきデータをバッファリングできる。

【0030】またログを応用すると、不必要なデータ転送の削減も達成できる。その例を次に示す。ここで、銀行の口座残高のデータベースが二重化されており、副ホスト上のデータベースはこのデータを一日に一回（例えば夜中）にしか参照しないような運用を想定する。このようなケースで、顧客が一日に複数回預金の出し入れを行った場合、各出し入れ時毎に副ホスト側のディスク制御装置に預金残高のデータが送られる。しかし、副ホスト側のデータベースは一日に一回しかデータを参照しないため、実際には最後のデータだけを送れば必要十分である。このような不必要なデータ転送の防止は、リモートコピーによるデータ二重化を間欠的に行うことにより実現できる。すなわちリモートコピーの中断と再開を一定時間毎に繰り返すのである。中断状態があまり長くなると、コピーするデータの情報を格納するログがオーバーフローする可能性があるため、主ホスト側のディスク制御装置のデータの更新頻度に応じて適当に調整する。勿論、副ホスト側のデータベースがデータを参照する場合は、リモートコピーを再開させ、コピー元とコピー先のボリュームの内容を同じにしておく。

【0031】さらに、リモートコピーの中断と再開の繰り返しと共に、ログの取り扱いも次のように変更する。リモートコピーの中断時間中は、更新されたデータの位置をログに登録するときに、既に同じ位置の更新がログに登録されていた場合には古い登録を抹消するのである。このようにすることにより、中断時間中に複数回更新されたデータは最後のデータだけが送信されるようになり、ディスク制御装置間での不必要なデータ転送が抑止できる。

【0032】最後に、本実施例では図5のように別のディスク制御装置7a、7b間でリモートコピーを行っている構成を説明している。しかしコピー先が同一ディスク制御装置の場合、すなわちボリューム二重化を行っている構成に関しても、図1に示す手順において、リモートコピーを二重化と読み替えることにより適用可能であり本発明の目的は達せられる。

【0033】

【発明の効果】本発明によれば、データをFTP等のファイル転送で送る代わりに、リモートコピー等ディスクシステムが自動的に事前にデータの二重化を行う方式を用いて、DB表データ等のデータをコピーしておき、データ転送時間を削減する方式において、コピー先のDB表データ等のデータが整合性の取れたデータであることを保証することが可能になる。これにより、データの転送時間を削減すると共に、基幹業務システム上でのデータ抽出処理を、データウェアハウスシステム側にオフロードでき、データ抽出処理時における基幹業務システムのレスポンス低下も回避できる。

【図面の簡単な説明】

【図 1】本発明における、整合の取れた DB 表をリモートコピーし、副システム側で参照する手順の一例を示す図である。

【図 2】ホストコンピュータに接続された単一のディスク制御装置の一構成例を示す図である。

【図 3】リモートコピーシステムの一構成例を示す図である。

【図 4】データベースシステムとディスク上の DB 表との関係の例を示す図である。

【図 5】ホストコンピュータに DBMS を備えている場合のリモートコピーシステムの一例を示す図である。

【図 6】本発明における、整合の取れた DB 表をリモートコピーし、副システム側で参照する手順の他の例を示す図である。

【図 7】ホストおよびディスク制御装置間をスイッチで接続した構成の一例を示す図である。

【図 8】リモートへコピーするデータの位置を示すログの一構成例を示す図である。

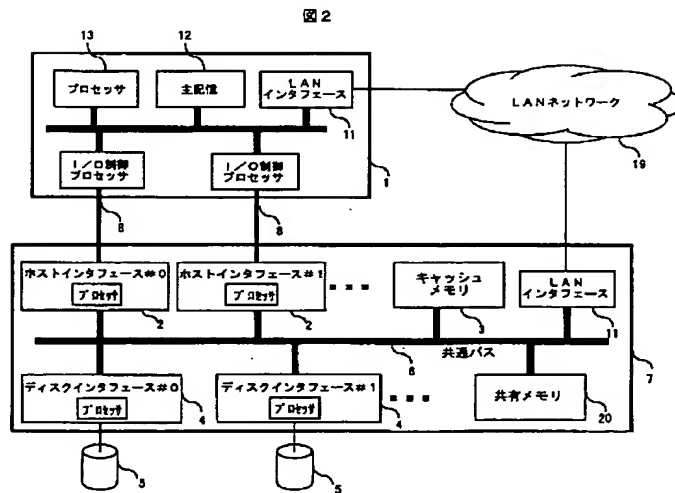
【符号の説明】

- 1、1a、1b ホストコンピュータ
2 ホストインタフェース

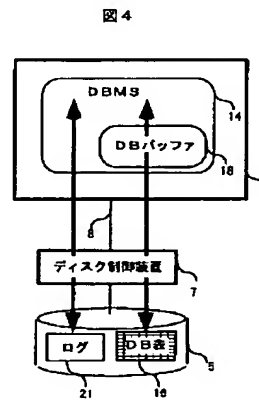
- * 3 キャッシュメモリ
4 ディスクインタフェース
5、5a、5b ディスク駆動装置
6 共通バス
7、7a、7b ディスク制御装置
8 チャンネルバス
9 主システム
10 副システム
11 LAN インタフェース
12 主記憶
13 プロセッサ
14、14a、14b データベース管理システム
15a、15b エージェント
16 データベース表
17 コピーされたデータベース表
18、18a、18b DB バッファ
19 LAN ネットワーク
20 共有メモリ
21 ログファイル
22 チャンネルバススイッチ

*

【図 2】



【図 4】



【図 8】

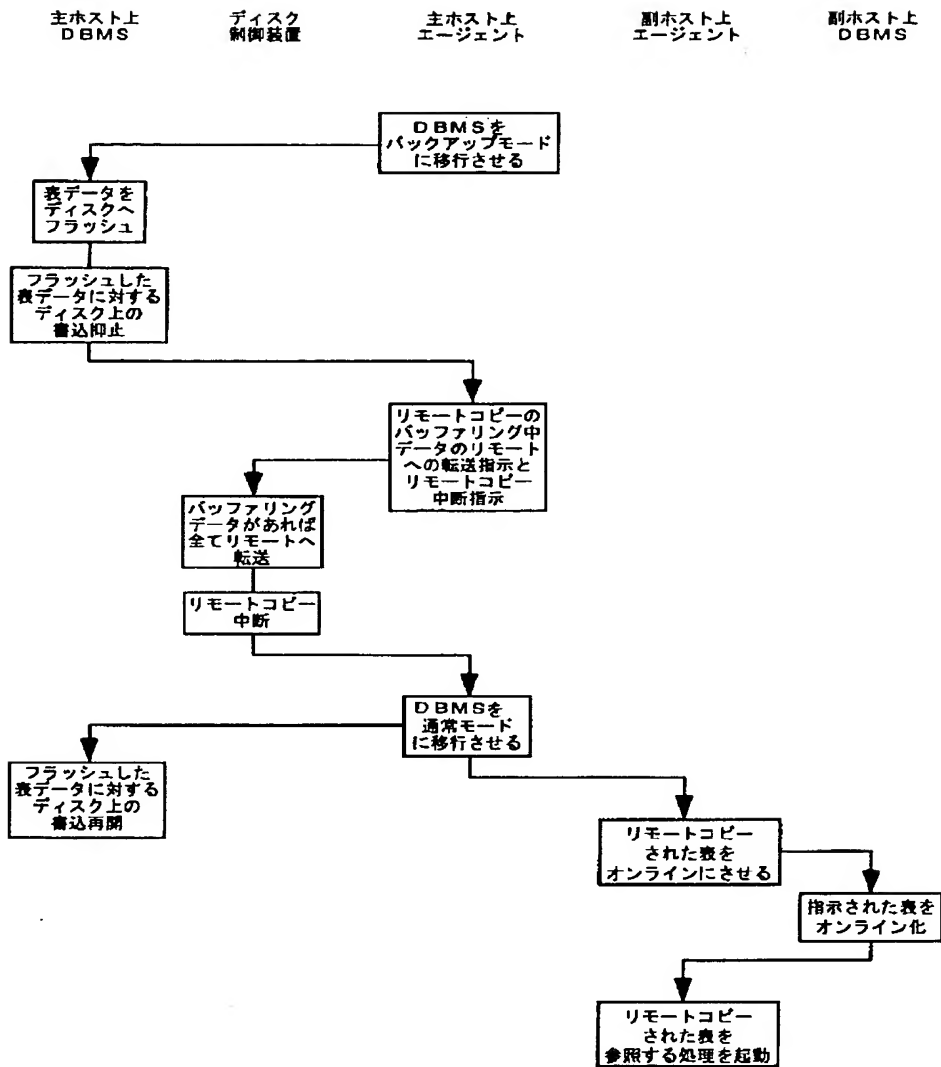
図 8

ボリウム番号	シリンダ番号	ヘッド番号	レコード番号
0	1	1	1
0	1	1	2
0	1	1	3
0	5	1	1
0	6	1	1
1	1	2	1
1	2	3	1
1	3	4	1

⋮

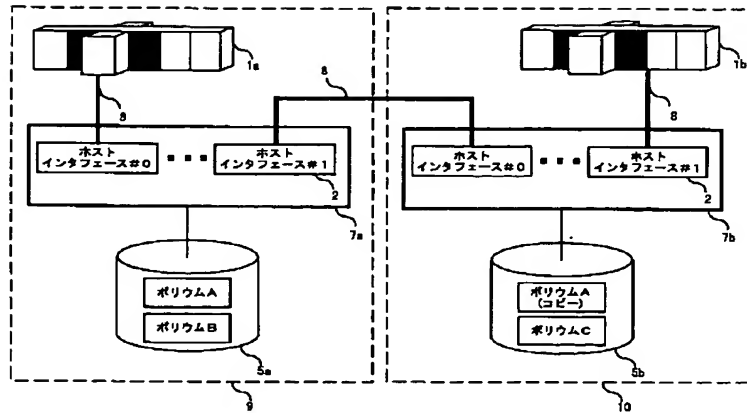
【図1】

図1



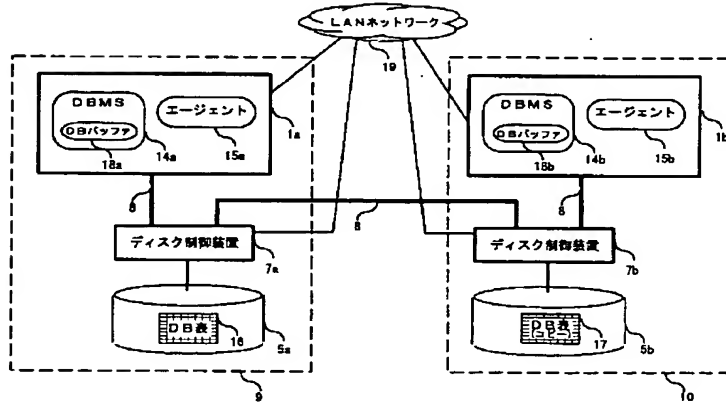
【図3】

図3



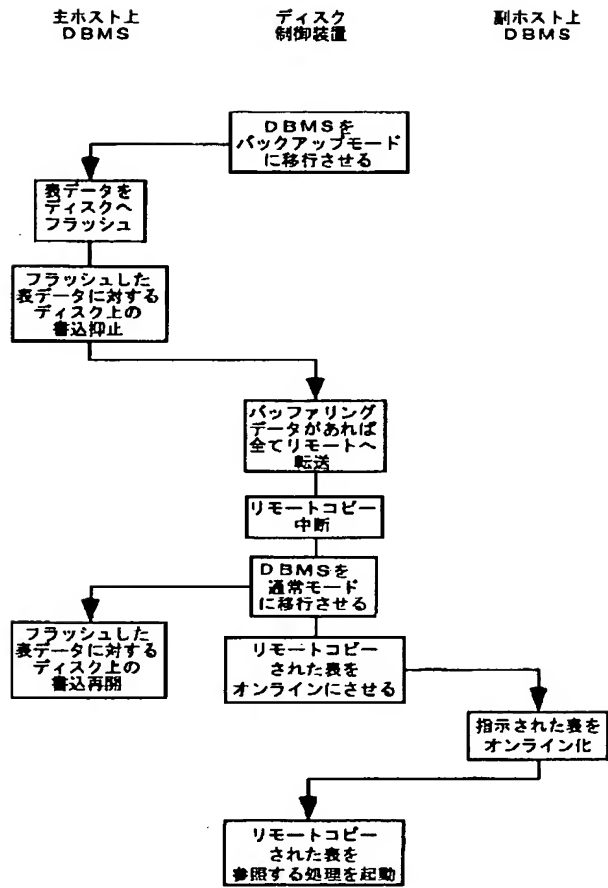
【図5】

図5



【図6】

図6



【図7】

図7

